

Quality assurance through laboratory reference materials (LRMs): dataset for non-certified elements (Al, Fe, Li, Mn and Pb) in estuarine sediment BCR-277R

C. Guitart^{1*}, C. López-Galindo¹, M. Rejas², F. Schultze³, A. Gamaza¹, J.L. Fernández-Turiel², V. Besada³, J. Benedicto¹

¹Instituto Español de Oceanografía (IEO), Centro Oceanográfico de Murcia. Varadero 1, 30740 San Pedro del Pinatar (Murcia), Spain

²labGEOTOP Facility, Laboratory of Elemental and Isotopic Geochemistry for Petrological Applications, Institute of Earth Sciences Jaume Almera (ICTJA-CSIC), Lluís Solé i Sabaris s/n, 08028 Barcelona, Spain

³Instituto Español de Oceanografía (IEO), Centro Oceanográfico de Vigo. Subida a Radio Faro 50, 36390 Vigo (Pontevedra), Spain

Author e-mail: carlos.guitart@mu.ieo.es Phone: +34 968179427; Fax: +34 968 184441

Abstract

Environmental and analytical laboratories have to perform quality assured measurements. However, this is limited due to the lack of appropriate matrix certified reference materials (CRMs), which ultimately frustrates the laboratory quality assurance (QA) process. Although the number of emerging and regulated environmental chemical contaminants is growing with subsequent societal and legal implications, the analytical scientific tools for both method and data validation are staying behind. With the aim to build on this issue, three specialized laboratories examined the determination of non-certified trace elements (Al, Fe, Li, Mn and Pb) in a commercial CRM and therefore, produced a so-called laboratory reference material (LRM). The total mass fractions and their associated uncertainties were estimated both using parametric and non-parametric methods. The results provided a coherent assessment for all except for Pb, in agreement with the results found by the CRM producer. We discuss here, the feasibility of the implementation of small-sized specialized laboratory comparisons to produce LRMs (from CRMs) for non-certified chemicals and strongly discourage attempts by a single laboratory.

Keywords: trace elements, marine, sediment, CRMs, LRMs, reference material, quality assurance

1. Introduction

Thousands of laboratory chemical measurements need to be made every day in many scientific disciplines (e.g. environment, biotechnology, food science, geochemical, etc.), guaranteeing reliable, comparable and quality assured analytical results. For the latter, national and international organizations are dedicated to contribute, primarily, through the so-called matrix reference materials (RMs), certified reference materials (CRMs) or standard reference materials (SRMs). Their production and certification follows internationally agreed standard protocols depicted by the International Organization for Standardization (*viz.* ISO guides), prior to sale and distribution. A view of this global market can be obtained visiting a web-based service assisting analytical laboratories in finding the CRMs they need (www.comar.bam.de). The goal of these matrix CRMs is to provide guidance (a *„measurement benchmark“*) to guarantee an up to standard individual analytical laboratory performance.

For example, in the marine pollution and environmental research field, few national metrology institutes (e.g. National Institute for Standards and Technology (NIST), USA) and specialized organizations, such as the International Atomic Energy Agency (IAEA, Austria) produce these matrix CRMs to determine trace metals and organic contaminants in marine organisms (e.g. mussel, tuna, shrimp, etc.) or sediments of diverse origin and level of contamination (e.g. harbors, estuaries, coastal, open sea, etc.) [1]. Consecutive matrix CRMs are produced or recertified due to the potential for deterioration for some matrices, such as the 25-year old serie CRM 278 (1988) - CRM 278R (1998) - ERM® 278 (2004) and ERM®-CE278k (2012), developed and produced by the BCR Office (Bureau Communautaire de Référence, European Commission, Belgium) and the EU-JRC-IRMM (European Commission-Joint Research Center-

Institute for Reference Materials and Measurements, Belgium) later on, for the trace metal determinations in the mussel matrix. This CRM serie is still being used to fulfill with the analytical data quality requirements for the laboratories involved in the marine Mussel Watch Monitoring Programs [2-4]. In the same way, the sediment matrix material BCR®-277R (2006) (the test material selected in this study) is the descendant of the original BCR-277 (1988) sold out by the EU-JRC-IRMM. Fundamentally, through CRMs, both analytical methods and measurements can be continuously quality assured and the chemical databases can be validated allowing for further science-based marine local, regional and global environmental assessments. Currently, the BCR®-277R it is called among others, to serve for the validation of the sediment contamination databases and therefore, to support the subsequent marine environmental assessments within the ongoing implementation of the European Union Marine Strategy Framework Directive [5].

For these reasons, the certification of the priority and regulated chemical targets in commercial matrix CRMs are essential in many research and applied fields. However, there are a number of reasons, including quality thresholds, experimental and organizational limitations, budget cuts, etc., which impede for their certification. In addition, there is a random combination of the required analytes (e.g. emerging or classical chemicals) with the appropriate matrices from different producers to some extent, which imply the use of 2 or 3 expensive CRMs by laboratories for their routine measurements. To mention few essential targets, for example, the mass fractions of major and minor matrix constituents in marine sediments, such as aluminum (Al) and lithium (Li), respectively, are important complementary elements used for the data normalization of trace metal distributions [6]. As well, either natural or anthropogenic lead (Pb) concentrations occurring in the marine environment are of major concern,

being the Pb levels and environmental monitoring regulated under diverse food or environmental legislation. More, iron (Fe) and manganese (Mn) are also two important environmental elements (i.e. micronutrients) usually monitored in marine sediments, whilst not classified as contaminants. Overall, these trace metals and the matrix major and minor components sums to an increasing number of anthropogenic emerging chemicals of concern (mostly organic contaminants) in different marine environmental compartments [7] that requires analytical validation. It should be pointed here, that this commercially CRMs contain most of these chemicals; hence, the majority of those are prepared from bulk samples directly collected in contemporary anthropogenic impacted marine environments [8].

Most individual specialized and research laboratories have the experience and capacity to contribute by determining these complementary or emerging chemical targets in CRMs, in line with the so-called LRMs [9], as they are both testing laboratories and end-users of CRMs. As well, their quality process needs to be fulfilled one way or another. In this work, we assess the lacking Al, Fe, Li, Mn and Pb (as a model compounds) in a commercial CRM, with a minimal number of highly specialized laboratories. We discuss from practical, statistical and long-term viewpoints the validity of a LRMs approach (originated from CRMs), as whether it could be reliable for analytical quality assurance purposes.

2. Materials and methods

A portion of the selected matrix material, an estuarine sediment certified reference material (BCR®-277R), was distributed among three laboratories (numbered 1, 2 and 3) to undertake the determination of Al, Fe, Li, Mn and Pb by using their own

analytical methodologies. In this work, the overall comparison for the LRMs approach followed the basic premises: a) a number of highly specialized laboratories participation (n=3, two marine analytical laboratories and one analytical laboratory specialized in rocks and soils); b) the determination of the analytical targets by different instrumental techniques; c) QA verified by CRMs performance for participant laboratories; and for the data evaluation, d) the use of both parametric and non-parametric statistics.

2.1. Sample pretreatment and digestion methods

The moisture content was calculated by laboratory 1 drying the whole material (2 g) in an oven at 40 °C during 12 hours, whilst laboratories 2 and 3 dried about 300 mg of the material for 24 hours in an oven set to 105°C (i.e., constant weight). All the laboratories corrected their results by the moisture content. Laboratory 1, performed the digestion of the sediment triplicates (about 100 mg each) using a combination of acids (2.5 mL HNO₃ + 5 mL HF + 2.5 mL HClO₄) in a closed-environment system (EVAPOCLEAN®, AHF analysentechnik) at 130 °C during 24 hours. After the evaporation process, the residue was collected and acidified by adding 1 mL of HNO₃ and transferred into high density polyethylene (HDPE) 100 mL volumetric flasks and stored until analysis. A microwave-assisted oven digestion procedure was used by laboratories 2 and 3 (CEM MARS Xpress and Milestone ETHOS Pro instruments, respectively), using a similar methodology. Basically, sediment triplicates (about 350 mg each) were weighted into the microwave vessels and prepared for digestion by adding a mixture of acids, 1 mL of aqua regia (HCl/HNO₃ 3:1) and 5 mL of HF. The oven program was set at 160°C using a 10 minutes ramp time and hold for 45 minutes for both laboratories. After the digestion step and the neutralization of the residual HF

with boric acid (4.7 g) the extract was adjusted to 50 mL in volumetric flasks with ultra-pure deionised water (Milli-Q, Millipore) ready for instrumental analysis.

2.2. Analytical instrumentation and QA

Laboratory 1 performed the analyses by using an inductively coupled plasma high resolution mass spectrometer (ICP-HRMS, Sector Magnetic Field Element XR, Thermo Scientific). Main ICP acquisition parameters were: sample gas (1.215 L min^{-1}) using and standard nebulizer (0.4 mL min^{-1}) and spray chamber (Helix Twinnabar, 6 mm) and Argon pressures (middle and maximum, 3.12 and 5.82 bar, respectively). The HRMS conditions were $4.0\text{e-}007$ mbar (high vacuum) and $2.5\text{e-}003$ mbar (fore vacuum) and maintained at -7941V (high voltage). The element isotopes ^7Li , ^{27}Al , ^{54}Fe , ^{55}Mn and ^{208}Pb were chosen in low resolution mode, including ^{27}Al in medium resolution, for quantification purposes and using ^{115}In (100 ng g^{-1}) as internal standard. The calibration standards were prepared from commercial certified solutions. Unielemental solutions were used for Li, Fe, Al and Mn ($1000 \text{ }\mu\text{g g}^{-1}$, matrix 2-5 % HNO_3 , VWR International LTD) and a multielemental solution for Pb (CCS-6, concentration $100 \text{ }\mu\text{g g}^{-1}$, matrix 7% HNO_3 , Inorganic Ventures). Calibration solutions ranged between $1 \text{ }\mu\text{g g}^{-1}$ to $100 \text{ }\mu\text{g g}^{-1}$ for Al and Fe, between 10 ng g^{-1} to 500 ng g^{-1} for Li and Mn and, finally 0.1 ng g^{-1} to 10 ng g^{-1} for Pb. The ICP-HRMS was optimized daily, followed by calibration and procedural blanks checks. The matrix materials GSJ rock (JA-2 and JB-3) and a QUASIMEME sample (marine sediment QTM087MS) were processed with the target sediment sample.

Laboratories 2 and 3 used atomic absorption spectrometry (AAS) for all the element determinations. Flame-AAS (Perkin Elmer AAnalyst 800 and Perkin Elmer Analyst

100, respectively) were used for Al, Li, Mn and Fe determinations, whilst Pb was determined by Zeeman graphite furnace-AAS (Perkin Elmer AAnalyst 800 and Perkin Elmer 4110 ZL, respectively), as well as Mn determination by laboratory 2. Both laboratories prepared calibration standards from commercial certified solutions (1000 $\mu\text{g L}^{-1}$ from Merck and Scharlau, respectively). Laboratory 2 working linear calibrations ranged up to 80 mg L^{-1} for Al, 40 mg L^{-1} for Fe, 2 mg L^{-1} for Li, 200 $\mu\text{g L}^{-1}$ for Mn and 400 $\mu\text{g L}^{-1}$ for Pb. Laboratory 3 working linear calibrations ranged up to 60 mg L^{-1} for Al, 6 mg L^{-1} for Fe and Li, 2 mg L^{-1} for Mn and 200 $\mu\text{g L}^{-1}$ for Pb. A KCl solution was added to the samples (0.1% KCl) by both laboratories for Al determinations with the N_2O_2 /acetylene flame in FAAS, whilst a matrix modifier (c.a. 2%) $\text{Mg}(\text{NO}_3)_2 + \text{NH}_4(\text{H}_2\text{PO}_4)$ was used during Pb determinations by Zeeman GFAAS. All atomic absorption measurements were performed at recommended wavelengths (i.e., Al-308.2 nm and Al-309.3 nm, Fe-302.1 nm and Fe-248.3 nm, for laboratory 2 and 3, respectively, and equal for Li-670.8 nm, Mn-279.5 nm and Pb-283.3 nm). An instrument sequence run was typically a sensitivity check, a calibration curve and both procedural blanks and matrix reference material checks (e.g., IAEA-433, IAEA-158 and QUASIMEME sediments). Trace and ultra-trace element grade reagents (e.g., MERCK Suprapur, PANREAC Hiperpur or FISHER Optima grade), cleaning procedures and all the precautions involved in trace metal determinations configured the standard operational procedures followed by laboratories. Further, since 1990, laboratories 2 and 3 participate satisfactorily in external analytical quality programs, such as the IAEA Worldwide interlaboratory comparison (ILCs) exercises [10,11] and QUASIMEME Laboratory Performance Studies (Quality Assurance of Information in Marine Environmental Monitoring in Europe), for trace element determinations in marine samples. Instead, laboratory 1 is involved in the international proficiency tests for

analytical geochemistry laboratories organized by the International Association of Geoanalysts (IAG). Analytical method descriptions can be found from authors elsewhere [12, 13].

2.3. Statistical assessments

The statistical assessment of the results was approached by using both parametric and non-parametric methods taking in consideration the minimal number of laboratories and datasets. Initially, the Shapiro-Wilk test (normality test, $\alpha=0.05$) was performed to contrast the null hypothesis of the pooled analytical datasets to fit into a normal probability distribution for each element provided by laboratories. A number of central tendency statistics, namely, arithmetic mean, 5% trimmed mean, median and central robust estimators (Huber's M-estimator, Tukey's biweight, Hampel's M-estimator), were calculated to both investigate and elucidate appropriate target values. Further, data dispersion estimators, namely, standard deviation (SD), interquartile range (IQR), coefficient of variation (CV) or residual standard deviation (RSD), maximum and minimum, and the 95 % confidence intervals for the mean (95% CI) were also calculated. The element recoveries in the commercial test material were calculated afterwards, based solely on parametric statistical results, to provide information on laboratories performance. After the initial data treatment, the final laboratory comparison results and their associated uncertainty were calculated based on the mean values of the replicate determinations by laboratories (eg. mean of laboratory means, etc.) rather than pooled, as different analytical methodologies were used. All the statistics, including exploratory analysis, Box-Whisker plots and statistical data typification (not shown) were carried out using the SPSS Statistical Software 17.0.

3. Results and discussion

3.1. Examination of the datasets

Figure 1 summarizes the analytical performance of each laboratory in terms of accuracy (% of recovery to the arithmetic mean determined in this study) for Al, Fe, Li, Mn and Pb. Overall, there was an agreement between the datasets produced for Fe, Li and Mn determinations, despite slightly dispersed results were observed for Al, and especially for Pb. All the combined datasets were assumed to belong to a normal probability distribution, even if the Shapiro-Wilk (SW) test was significant for Li and Pb datasets (p-value 0.003 and 0.000, respectively). The reasons for these non-normal distributed datasets were caused by statistical outliers (data typification values of -2.5 and -1.3/-2.1, for Li and Pb, respectively), therefore, excluding these values the SW test was not significant (p-value 0.220 and 0.525, respectively). Nevertheless, these outliers were not removed for further statistical calculations, as we considered these were correct values reflecting the measurement variability encountered even by quality assured and specialized laboratories. Alternatively, a 5% trimmed mean and a number of central robust estimators were calculated based on non-parametric statistics. The robust statistics complemented the assessment process of the target analytes by contrasting the degree of divergence between both statistical approaches, and therefore providing a better judgment. Table 1 and 2 resume the parametric and non-parametric statistical estimations.

Figure 1 shows an excellent agreement between laboratories 1, 2 and 3 for Fe (96%, 98%, 106%, respectively), Li (102%, 104%, 94%) and Mn (108%, 96%, 97%).

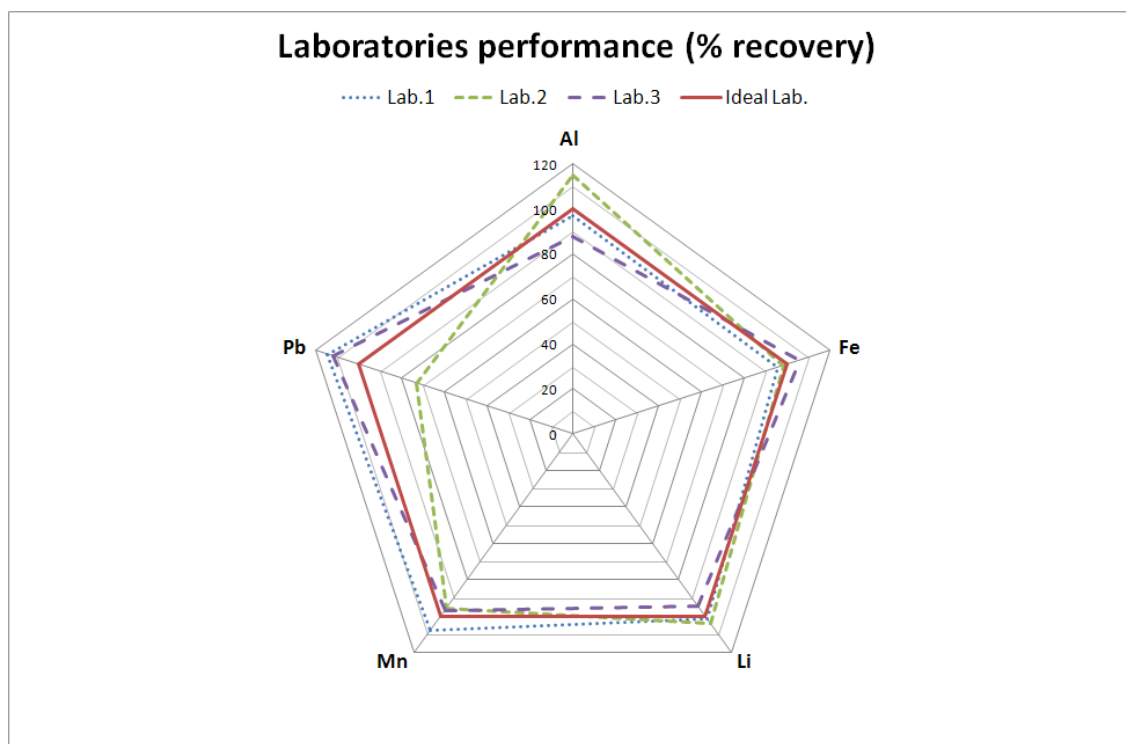


Figure 1. Laboratories performance in terms of accuracy (% of recovery) calculated using the arithmetic mean concentration values determined in this study (a 100% recovery is given for an ideal laboratory).

Both individual ICP-HRMS and AAS instrumental determinations for these elements gave CVs lower than a 5%, except for Li determination by AAS for laboratory 3 which increased to a 9%. The pooled CV for the whole datasets was 5%, 6% and 6% for Fe, Li and Mn, respectively (Table 1). Consequently, the arithmetic mean, the 5% trimmed mean, the median and all the robust central estimators (i.e., Huber's M-estimator, Tukey's biweight and Hampel's M-estimator) largely coincide (Table 1 and 2). Accordingly, the concentration ranges obtained with 1 SD of the mean, the median and their associated IQR, or the mean 95% CI (Table 1 and 2), encompass the statistical results.

The individual laboratory Al determinations gave CVs lower than 3% but an overall CV of 12% between laboratories, thus the data ranged from 61.5 g kg⁻¹ (minimum) to 83.6 g kg⁻¹ (maximum) (Table 1), and recoveries were 97%, 115% and 88% for laboratory 1, 2 and 3, respectively (Figure 1). Correspondingly, all the

statistical results provided also similar central tendency estimations, although both the SD and the IQR increased, reflecting more dispersed results. The ICP-HRMS determinations for Al (laboratory 1) were the most precise ($70.3 \pm 1.4 \text{ g kg}^{-1}$) and laboratory 3 provided the lowest accurate concentration ($63.5 \pm 1.8 \text{ g kg}^{-1}$). Therefore, determinations by AAS gave both higher and lower values than those performed by ICP-HRMS. The Al determination variability by F-AAS (laboratory 2 and 3), can be attributed to a number of causes, such as the N_2O_2 /acetylene flame stability, accurate sensitivity checks, or changing standard and sample solution conditions during the analytical measurements.

Table 1. Statistical parametric results for the target major, minor and trace elements in the matrix material BCR®-277R.

Element	Mean	Standard deviation (SD)	Coefficient of Variation-RSD (%)	Confidence Interval (CI) $\pm 95\%$	Minimum value	Maximum value
Al (g kg^{-1})	72.4	8.8	12	65.6 - 79.1	61.5	83.6
Fe (g kg^{-1})	43.5	2.3	5	41.7 - 45.2	40.9	47.4
Li (mg kg^{-1})	55.4	3.4	6	52.9 - 58.0	47.2	58.1
Mn (mg kg^{-1})	930	57	6	886 - 975	875	1033
Pb (mg kg^{-1})	29.8	8.4	28	23.3 - 36.2	12.0	35.0

Table 2. Non-parametric and robust statistical results for the target major, minor and trace elements in the matrix material BCR®-277R.

Element	Trimmed Mean (5%)	Median	Interquartile range (IQR)	Central Robust Estimators		
				Huber's M-estimator	Tukey's biweight	Hampel's M-estimator
Al (g kg^{-1})	72.3	70.0	18.7	71.0	71.5	71.9
Fe (g kg^{-1})	43.4	42.6	3.5	42.8	42.4	42.7
Li (mg kg^{-1})	55.8	56.9	2.8	56.3	56.5	56.4
Mn (mg kg^{-1})	928	917	105	919	919	925
Pb (mg kg^{-1})	30.5	33.7	8.6	33.6	33.8	33.9

From a practical perspective, the determinations of Al, Fe, Li and Mn are examples of 'easy-to-do' elements provided the laboratories possess the required

analytical expertise as demonstrated in this work, thus a consistent analytical determination between laboratories was achieved. Table 3 shows the final estimated mass fractions and their associated uncertainties using JCGM 100:2008 guidelines [14]. The way forward to complete the LRM process should be to establish the traceability of Al, Fe, Li and Mn to a similar analyte-matrix CRM where these elements are fully certified (if it exist), as well as to be able to demonstrate QA performance by the involved laboratories in the long-term (Table 3). Thus, analytical laboratories might perform small-sized comparisons for non-certified analytes in their CRMs in order to use them then, as LRMs. The advantage would be that all the target chemicals to be determined would be present in the desired matrix material saving time and money (either certified or assessed), rather than to use multiple CRMs. More, the use of CRMs for this purpose benefits as the material has been treated and prepared accordingly (i.e. homogenization, sterilization, sieved, etc.). We show it is possible to produce a LRM based on a minimal number of laboratories, however, if there is any analyte data disagreement beyond known or required analytical standards this approach should be dismissed. This was observed for Pb and is discussed in the next section.

Table 3. Analytical results for the small-sized laboratory comparison and z-scores obtained by the organizing laboratory (number 3) for the QUASIMEME Proficiency Tests (latest data available).

Sediment material	Al (g kg⁻¹)	Fe (g kg⁻¹)	Li (mg kg⁻¹)	Mn (mg kg⁻¹)	Pb (mg kg⁻¹)
BCR-277R					
<i>This study</i> ¹	72.4	43.5	55.4	930	29.8
<i>Combined SD</i> ²	1.2	1.1	2.7	23	6.9
<i>Expanded Uncertainty</i> ³	2.5	2.2	5.3	46	13.7
QUASIMEME Z-scores for sediment materials (Laboratory 3)					
2011	-1.1	0.0	-0.4	-1.0	0.1
2011 (2 nd)	-0.3	0.5	0.3	0.3	0.4
2012	-0.3	0.2	-0.5	-0.1	-1.1

¹Mean of laboratories means; ²Square root of combined (best estimated) variances from laboratories;

³Combined uncertainty with a coverage factor k=2

3.2. Lead troubleshooting

The Pb concentrations were determined in the BCR®-277R material in close agreement by laboratory 1 and 3 using ICP-HRMS and ZGF-AAS, respectively ($34.1 \pm 0.5 \text{ mg kg}^{-1}$ and $33.4 \pm 0.4 \text{ mg kg}^{-1}$), and with different pre-treatment and digestion methods as well. Surprisingly, laboratory number 2 provided highly dispersed results ($21.8 \pm 11.9 \text{ mg kg}^{-1}$), rising the CV of the whole comparison dataset up to a 28%. As mentioned in section 2.3., the datasets from all the participants (provided by specialized and quality assured laboratories) were trusted and therefore, these Pb results were not discarded. The recoveries were 115%, 73% and 112% for laboratory 1, 2 and 3, respectively (Figure 1). The laboratories dataset mean (29.8 mg kg^{-1}) and the median (33.7 mg kg^{-1}) show the highest statistical discrepancies in this work. For that reason, the parametric (i.e., arithmetic mean and 5% trimmed mean) and the non-parametric results (i.e., median and central robust estimators) group well around their respective statistics (Table 1 and 2). The overall SD obtained for the mean ($\pm 8.4 \text{ mg kg}^{-1}$) is about a 50% higher than the usual standard deviations given in similar CRMs, but comparable to the total errors assigned in external comparison studies. In a similar way, the IQR ($\pm 8.6 \text{ mg kg}^{-1}$) rises up to one third of the median value (Table 2) as well, despite suggesting similar concentration ranges.

It is complex and would require further laboratory investigations down to the matrix characteristics of this sediment material (e.g. mineral composition, particle size distribution, organic matter type and content, etc.), to ascertain why the Pb analytical datasets between specialized laboratories have shown this unexpected variability. In any case, it can be stated straightforward, that any of the specialized laboratories may be biased when conflicting results appear. It should be particularly mentioned here, that the

report describing the certification process of the sediment material BCR®-277R by the EU-JRC-IRMM, states: *“the data obtained in the characterisation was not suitable to certify Pb in this material”* in Held et al., 2006 [15]. Therefore, our results from a small-sized comparison coincides with the above mentioned report and clearly point to the potential for gross errors if a single laboratory attempt to perform the analytical determination of non-certified chemicals in CRMs (or another matrix material such as an *“in-house”* material) for their QA purposes. In other words, internal laboratory checks (*viz.* internal reference materials IRMs) should be disregarded as a valid approach for laboratory QA in the medium and long-term. We demonstrate in this work, a valid LRMs approach when all the datasets from a small-sized laboratory comparison fully agree and then, the process should be completed during the routine measurements through establishing the traceability to another CRM if it exists. Obviously, for emerging and new potential contaminants this approach could be a starting point, thus traceability could not be establish until proper CRMs would be available.

4. Conclusion

The scientific approach in this study based on small-sized laboratory comparisons to perform the analytical determinations for non-certified target chemicals in CRMs can be helpful to generate temporary LRMs whilst awaiting for optimal CRMs. When there is full analytical and statistical agreement between specialized laboratories the traceability of the produced LRM should be established, if possible, to another CRM. We demonstrate that minimal discrepancies between datasets point to potential issues with the analyte-matrix material and the results should be discarded. In this way, this LRMs (or laboratory fit-for-purpose CRMs) could be used to monitor the

QA performance saving time and money, but provided the laboratory can demonstrate long-term consistency of their analytical measurements. We also demonstrated that the LRMs approach should not be undertaken by a single laboratory. Errors still always around, even in high specialized and quality assured laboratories, and without doubt the CRMs play a fundamental role.

Acknowledgements

Analyses of ICTJA-CSIC were carried out in the Geochemistry Facility of labGEOTOP, infrastructure co-funded by ERDF-EU (Ref. CSIC08-4E-001). The IEO acknowledges the funding received through the Spanish Ministry of Agriculture, Food and Environment (Encomienda Gestión IEO-MAGRAMA 2010-2012). We also acknowledge the useful comments received by anonymous reviewers.

References

- [1] S.J. De Mora, J. Villeneuve, E. Wyse. 2007. An overview of laboratory performance studies conducted by the IAEA for marine pollution studies of metals and organic contaminants. *Accred. Qual. Assur.* 12 (2007) 587-592.

- [2] K.L. Kimbrough, W.E. Johnson, G. Lauenstein, J.D. Christensen, D.A. Apeti. *An Assessment of Two Decades of Contaminant Monitoring in the Nation's Coastal Zone.* Silver Spring, MD. NOAA Technical Memorandum NOS NCCOS 74. 105 pp, 2008.

- [3] C. Guitart, A. Hernández-del-Valle, J.M. Marin, J. Benedicto. Tracking temporal trend breaks of anthropogenic change in Mussel Watch (MW) databases. *Environ. Sci. Technol.*, 46 (2012) 11515-11523.
- [4] V. Besada, J.L. Sericano, F. Shultze. An assessment of two decades of tracemetals monitoring in wild mussels from the Northwest Atlantic and Cantabrian coastal areas of Spain, 1991-2011. *Environ. Int.*, 71 (2014) 1-12.
- [5] EU MSFD, 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive).
- [6] M.O. Angelidis, O. Radakovitch, A. Veron, M. Aloupi, S. Heussner, B. Price. Anthropogenic metal contamination and sapropel imprints in deep Mediterranean sediments. *Mar. Poll. Bull.*, 62 (2011) 1041-1052.
- [7] S. Bricker, G. Lauenstein, K. Maruya. NOAA's Mussel Watch Program: Incorporating contaminants of emerging concern (CECs) into a long-term monitoring program. *Special Issue Mar. Poll. Bull.: NOAA's Mussel Watch Program* 81 (2014) 289-290.
- [8] A.M. Peck, J.R. Kucklick, M.M. Schantz. Synthetic musk fragrances in environmental Standard Reference Materials. *Anal. Bioanal. Chem.*, 387 (2007) 2381-2388.

- [9] P. Quevauviller and E.A. Maier. Interlaboratory studies and certified reference materials for environmental analysis. The BCR approach. Techniques and instrumentation in analytical chemistry Vol. 22. Elsevier Science, Amsterdam, The Netherlands, 1999.
- [10] E.J. Wyse, S. Azemard, S.J. De Mora. Report on the world-wide intercomparison exercise for the determination of trace elements and methylmercury in marine sediment IAEA-433. IAEA/AL/147, IAEA/MEL/75, IAEA, Vienna, Austria, 2004.
- [11] M.J. Campbell, S. Azemard, J. Oh. Report on the world-wide intercomparison exercise on the determination of trace elements in IAEA-158 marine sediment. IAEA/AL/187, IAEA/MEL/82, IAEA, Vienna, Austria, 2008.
- [12] J. Benedicto, C. Martínez-Gómez, J. Guerrero, A. Jornet, C. Rodriguez. Metal contamination in Portman Bay (Murcia, SE Spain) 15 years after the cessation of mining activities. *Cienc. Mar.* 34 (2008) 1-10.
- [13] C. Quelle, V. Besada, J.M. Andrade, N. Gutierrez, F. Schultze, J. Gago, J.J. González. Chemometric tools to evaluate the spatial distribution of trace metals in surface sediments of two Spanish rías. *Talanta* 87 (2011) 197-209.
- [14] JCGM 100:2008. Evaluation of measurement data ó Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology-September 2008.
http://www.bipm.org/utis/common/documents/jcgm/JCGM_100_2008_E.pdf

[15] A. Held, G.N. Kramer, A. Lamberty, P. Robouch, U. Wätjen. The certification of the mass fractions of As, Cd, Co, Cr, Cu, Fe, Hg, Mn, Ni, Sb, Sc, Se, Sn, Th, Tl, U, V and Zn in three sediments BCR-277R estuarine sediment, BCR-280R lake sediment and BCR-320R channel sediment. European Commission, Directorate-General Joint Research Centre, Institute for Reference Materials and Measurements (IRMM, Belgium), 2006. https://irmm.jrc.ec.europa.eu/refmat_pdf/BCR-320r_report.pdf